Null models for dynamic centrality in temporal networks

Tim Rogers

Centre for Networks and Collective Behaviour,
Department of Mathematical Sciences, University of Bath,
Claverton Down, Bath, BA2 7AY, UK

t.c.rogers@bath.ac.uk

Dynamic centrality metrics provide a quantitative assessment of the strength of communication between nodes in temporal networks, as well as the overall capacity of the network for the efficient transmission of information. In this article the behaviours of two variants of the 'communicability' metric are examined in simple null models of uncorrelated temporal networks. Analysis of the long-time behaviour of the null models reveals a simple trade-off in the role of the parameters of the metric, suggesting methods to calibrate parameters and to adapt to temporal variations in the network properties. The null models introduced address two main classes of temporal networks (contact sequences and interval graphs), and their predictions are compared and contrasted with results coming from real-world telecommunications data.

Keywords: Temporal networks; centrality; graph metrics; null models.

1 Introduction

Understanding the relationship between the structure of a network and the efficiency with which information may be disseminated across it holds considerable scientific interest, as well as being of strategic value to government and commerce. In particular, the rise of online social networking has prompted questions about its potential as a method of mass-communication, for example in relation to emergency management [1]. To arrive at quantitative recommendations often requires high-dimensional network data to be analysed and distilled into summary statistics (metrics) which capture relevant properties of the network under study [2]. One of the earliest and most influential network metrics is *Katz centrality* [3], which is used as a indicator of the 'importance' of a given node in the network. Katz centrality is based on a weighted count of the number of walks between nodes in the network, on the premise that the capacity of one node to transmit information to another is related to the number and length of possible routes between them.

Although it is undoubtedly useful and has been widely applied, Katz centrality belongs to a class of network metrics which ignore a very important feature of real-world networks: temporal

structure. Very few applications involve a single network frozen in time; it is much more likely that some aspect of the interaction structure will be in flux. As such, it is now recognised that so-called temporal networks will play an increasingly important role in network analysis, particularly in relation to the live monitoring of complex systems such as online social networks. According to one popular classification scheme [4], there are two principle classes of temporal network: contact sequences, describing systems in which instantaneous messages are sent between nodes, and interval graphs, where links between nodes have a finite duration. Some traditional (static) network concepts and metrics have natural analogues in these paradigms, whilst other aspects of study must be completely overhauled.

Several authors have suggested new notions of centrality for temporal networks [5, 6, 7, 8, 9, 10]. A notable contribution comes from Grindrod *et al* [11, 12, 13], who have proposed generalisations of Katz centrality to temporal networks by observing that information flow must obey the arrow of time. The *communicability* metrics count weighted, time-respecting [4], walks between nodes. As well as weighting walks according to their length in the network, it is often necessary to consider their length in time; for many applications, communication that happened long in the past is less relevant than more recent contact. This consideration is made explicit in the communicability metrics by introducing a second parameter to downweight walks according to how long ago they started.

The *running communicability* metric is defined in [12] for temporal networks that can be written in the form of a sequence of adjacency matrices $\{A_1, A_2, \ldots\}$. In this scheme the matrix $\mathcal{S}$ is updated via the recursion equation

$$\mathcal{S}_{n+1} = (I + e^{-b(t_n - t_{n-1})}\mathcal{S}_n)(I - aA_n)^{-1} - I\,, \tag{1}$$

where a and b are the edge attenuation and time downweighting parameters and t_n is the timestamp corresponding to adjacency matrix A_n. The entry $\mathcal{S}_{ij}$ gives a weighted count of the number of walks from i to j, where a walk of length l that started t units of time in the past is given weight $a^l e^{-bt}$. While immediately well-suited to the analysis of contact sequences, further generalisation is required to define a metric suitable for interval graphs, in which the adjacency matrix is given as a function of time $A(t)$. Taking a limit of small timesteps in (1) results in version of the metric defined in continuous time. Introduced in [13], the *dynamic communicability* matrix $S(t)$, evolves in time according to

$$S'(t) = -bS(t) - \big(I + S(t)\big)\log\big(I - aA(t)\big)\,. \tag{2}$$

Between them, the metrics defined in (1) and (2) are applicable to a wide class of temporal network data. Through the choice of parameters a and b, the deterioration of information with distance and time can be accentuated. On first inspection it seems a and b are to be chosen arbitrarily; similar to Katz centrality in static networks, the only constraint is that the eigenvalues of aA_n or $aA(t)$ should lie within the unit disc.

In this article we will gain a deeper understanding of the communicability metrics by examining their behaviour in simple null models of temporal networks. The practical benefits of this analysis are twofold: (i) highlighting important differences between real-world temporal networks and what would be expected from random chance, and (ii) revealing a simple trade-off between the parameters a and b, which removes some of the arbitrary nature of their definition

and suggests a method to adjust their calibration in response to changes in the network. We will consider separately the cases of contact sequences and interval graphs in Sections 2 and 3, with example applications to email and telephone networks.

2 Directed contact sequences

2.1 Communicability in contact sequences

Consider a temporal network defined by a list of instantaneous directed contacts. The n^{th} such 'contact' is written as a triple (i_n, j_n, t_n) denoting a message sent from node i_n to node j_n at time t_n. Typical examples of such networks include email traffic and text messaging. The ordering of contacts is assumed to follow the arrow of time, so that $t_n \leq t_{n+1}$, although it should be pointed out that his ordering is not unique when there are multiple messages with the same time stamp.

As discussed in the introduction, the running communicability matrix $\mathcal{S}$ defined in (1) provides summary information about the flow of information in the network. For the specific case of directed contact networks, the update rule (1) may be simplified by treating messages individually in turn, without performing any prior time-aggregation. The adjacency matrix associated to a single message $i_n \to j_n$ has only one non-zero entry:

$$\big[A_n\big]_{ij} = \delta_{i,i_n}\delta_{j,j_n}\,. \tag{3}$$

It is straightforward to check that $(I - aA_n)^{-1} = I + aA$, which can be inserted into equation (1) to obtain

$$\big[\mathcal{S}_{n+1}\big]_{ij} = e^{-b\Delta_n}\big[\mathcal{S}_n\big]_{ij} + a\delta_{i,i_n}\delta_{j,j_n} + ae^{-b\Delta_n}\big[\mathcal{S}_n\big]_{i\,i_n}\delta_{j,j_n}\,, \tag{4}$$

where $\Delta_n = t_n - t_{n-1}$ is the waiting time before the arrival of message n. Explained in words: when message n arrives we (i) downweight each entry of $\mathcal{S}$ by a factor of $e^{-\Delta_n}$, then (ii) add a contribution of a to entry $\mathcal{S}_{i_n j_n}$ and (iii) add $aS_{i\,i_n}$ to $\mathcal{S}_{i\,j_n}$ for each i. These three steps correspond to downweighting by time elapsed, adding the new length 1 walk from i_n to j_n, and lastly adding the other walks to j_n from any starting point which use the new step $i_n \to j_n$.

It is worth pausing for a moment to discuss the treatment of messages with identical timestamps. Suppose two contacts (i_n, j_n, t_n) and $(i_{n+1}, j_{n+1}, t_{n+1})$ occur at the same instant in time, that is, $t_{n+1} = t_n$. Does their order in the contact sequence affect the computed values of communicability? Iterating equation (1), we see that after both messages have been sent, the communicability matrix is

$$\mathcal{S}_{n+2} = (I + e^{-b\Delta_n}\mathcal{S}_n)(1 - aA_n)^{-1}(1 - aA_{n+1})^{-1} - I\,. \tag{5}$$

The adjacency matrices are $[A_n]_{ij} = \delta_{i,i_n}\delta_{j,j_n}$ and $[A_{n+1}]_{ij} = \delta_{i,i_{n+1}}\delta_{j,j_{n+1}}$, so we can compute

$$\begin{aligned}(I - aA_n)^{-1}(I - aA_{n+1})^{-1} &= (I + aA_n)(I + aA_{n+1}) \\ &= I + a(A_n + A_{n+1}) + a^2 A_n A_{n+1}\,.\end{aligned} \tag{6}$$

The first two terms are clearly symmetric in n and $n+1$, so we must only check

$$\big[A_n A_{n+1}\big]_{ij} = \sum_k \delta_{i,i_n}\delta_{k,j_n}\delta_{k,i_{n+1}}\delta_{j,j_{n+1}} = \delta_{i,i_n}\delta_{j_n,i_{n+1}}\delta_{j,j_{n+1}}\,. \tag{7}$$

Thus $A_n A_{n+1}$ is empty unless $j_n = i_{n+1}$, in which case it has a one in row i_n and column j_{n+1}. This describes a situation in which the receiver of message n happens to also be sending another message to a third party at precisely the same time. For many systems this is not expected behaviour and, provided that the time resolution of the data is sufficiently fine, it may be reasonable to assume that events of this type are sufficiently rare that they can be ignored. Moreover, messages with multiple recipients (as often occur in email networks) can safely be treated as a collection of individual messages with the same timestamp; since each of these messages has the same sender the above situation is always avoided.

2.2 Null model

Historically, when dealing with static networks, sparse random graphs of the Erdös-Renyi type have been viewed as the canonical choice of null model. The edges of these networks are independently randomly assigned, thus any features of a real-world network not present in the random null model must be a consequence of correlation between edges (possible induced by some underlying confounding variable). Moving to a time-dependent setting, it is sensible to retain this independence in the location of edges, and to make the additional requirement that edges should be uncorrelated in time also (i.e. the model has no memory). These constraints uniquely specify the null model: the senders and receivers of messages are chosen uniformly at random, while the contact times obey a Poisson process with uniform rate. The model is parametrised by the size of the network N, and the average number of contacts per unit time μ.

How does communicability evolve under this model? Since the contacts of the null model are uncorrelated, the statistics of $\mathcal{S}$ are invariant under permutations of the rows and columns. To characterise the expected flow of information in the null model it is therefore sufficient to focus on a typical row or column sum. The *dynamic receive* score of a node i is calculated by summing the columns of (4), equivalent to counting weighted walks ending at i. For a general directed contact network we may use equation (4) to deduce

$$\begin{aligned}
\big[\boldsymbol{r}_{n+1}\big]_i &= \sum_j \big[\mathcal{S}_{n+1}\big]_{ji} \\
&= e^{-b\Delta_n}\sum_j \big[\mathcal{S}_n\big]_{ji} + a\delta_{i,j_n} + e^{-b\Delta_n}\sum_j \big[\mathcal{S}_n\big]_{j\,i_n}\delta_{i,j_n} \\
&= e^{-b\Delta_n}\big[\boldsymbol{r}_n\big]_i + a\Big(1 + e^{-b\Delta_n}\big[\boldsymbol{r}_n\big]_{i_n}\Big)\delta_{i,j_n}\,.
\end{aligned} \tag{8}$$

For the null model, we will study the expected dynamic receive $\rho_n = \mathbb{E}([\boldsymbol{r}_n]_i)$, where the average is taken over all possible realisations of the network. Note that, since all nodes are equivalent in the null model, the definition of ρ_n is invariant under different choices of i. There

are two sources of randomness in the model: the choice of the sending and receiving nodes and the time of the messages. Applying the expectation to (8) we obtain

$$\rho_{n+1} = \mathbb{E}\Big(e^{-b\Delta_n}\big[\boldsymbol{r}_n\big]_i\Big) + a\mathbb{E}\Big(\delta_{i,j_n}\Big) + a\mathbb{E}\Big(e^{-b\Delta_n}\big[\boldsymbol{r}_n\big]_{i_n}\delta_{i,j_n}\Big)\,. \tag{9}$$

To compute the averages here we first collect some facts about the null model. First note that the variables i_n, j_n and Δ_n are independent up to the constraint that $i_n \neq j_n$. Second, the equivalence of nodes implies that $\mathbb{E}([\boldsymbol{r}_n]_{i_n}) = \rho_n$, and $\mathbb{E}(\delta_{i,j_n}) = 1/N$. Finally, since the contact times obey a Poisson process with rate μ, we know that the waiting time Δ_n is exponentially distributed with rate μ. Combining these observations, we obtain the recursion relation

$$\begin{aligned}\rho_{n+1} &= \int_0^\infty \mu e^{-\mu\Delta}\left(e^{-b\Delta}\rho_n + \frac{a}{N}(1+e^{-b\Delta}\rho_n)\right)\,d\Delta \\ &= \frac{a}{N} + \left(1+\frac{a}{N}\right)\frac{\mu}{\mu+b}\,\rho_n\end{aligned} \tag{10}$$

The integral here comes from averaging over the exponentially distributed waiting time Δ_n, i.e. $\mathbb{P}(\Delta_n > \Delta) = e^{-\mu\Delta}$.

Depending on the relative values of the parameters, the sequence $\{\rho_n\}_{n\in\mathbb{N}}$ may converge, or may grow exponentially. For convergence we require

$$\left(1+\frac{a}{N}\right)\frac{\mu}{\mu+b} < 1\,, \tag{11}$$

or equivalently that $b/a > \mu/N$. There is thus a simple trade off between edge attenuation (a) downweighting in time (b) in terms of the number of messages sent per node per unit time. Loosely speaking, we must "forget" about old walks as quickly as new ones are added. The general solution of (10) is

$$\rho_n = \frac{b+\mu}{bN/a-\mu}\left(1-\left(\frac{\mu(1+a/N)}{\mu+b}\right)^n\right)\,. \tag{12}$$

In particular, provided $b/a > \mu/N$, we have

$$\lim_{n\to\infty}\rho_n = \rho^* = \frac{b+\mu}{bN/a-\mu} \approx \frac{\mu/N}{b/a-\mu/N}\,, \tag{13}$$

where we have assumed that N and μ are both large. We thus observe that, for mean communicability at least, the ratios b/a and μ/N determine the long-term behaviour of the system. We can exploit this relationship to calibrate the parameters in order to make meaningful comparisons between different systems or datasets. One obvious choice is to ask that $\rho^* = 1$ as $n \to \infty$, which determines the calibration

$$b \approx \frac{2a\mu}{N}\,. \tag{14}$$

For a given data set, one can easily compute the mean gap μ^{-1} between messages, then equation (14) gives a sensible calibration of b in terms of the edge attenuation parameter a. Alternatively, for live network monitoring the mean duration and gap can be calculated as rolling averages. This method has the considerable advantage of not being "stuck" with inappropriate parameters if the temporal characteristics of the network change (or, indeed, if the size of the network changes).

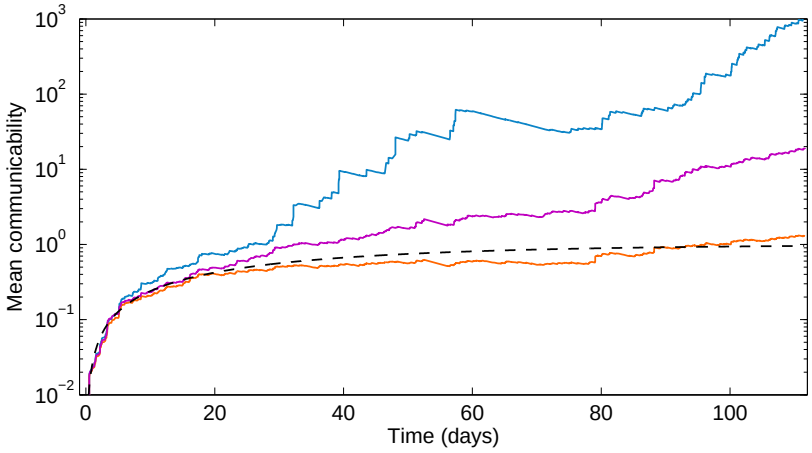

Figure 1: (Colour online) Growth of communicability in the Kiel email network with edge attenuation parameter $a = 0.5$. Dashed (black) line: null model prediction given in equation (12). Upper (blue) line: empirical average dynamic receive score in the Kiel network, computed using equation (15). Lower (orange) line: average dynamic receive with randomised message senders. Middle (magenta) line: average dynamic receive with partially randomised sender-receiver pairs, see the main text for details. For the randomised data the figure shows a single realisation from 1000 samples that were simulated – ensemble standard deviation was found to be negligible on this scale.

2.3 Application to email network

We move on now to compare the predictions of the null model with communicability data coming from a simple test case. The Kiel student email network [14] is an anonymised list of all emails sent to or from student accounts over a 112 day period; removing external emails and inactive nodes yields a contact list of 2011 messages, sent between $N = 328$ nodes. The mean waiting time between messages is $\mu^{-1} = 4812s$ (standard deviation $11053s$).

Starting from $\mathcal{S}_0 = 0$, applying the iterative procedure (4) creates a time series for communicability. To compare with the null model, we compute the empirical average of $\boldsymbol{r}_n$, which is obtained from $\mathcal{S}$ via

$$\widetilde{\rho} = \frac{1}{N} \sum_i \left[\boldsymbol{r}_n\right]_i = \frac{1}{N} \sum_{i,j} \left[\mathcal{S}_n\right]_{ij} . \tag{15}$$

In Figure 1 the null model (dashed black line) is compared with this empirical average (solid blue line) for $a = 0.5$ and b determined by the calibration (14). Although both time series agree initially, a profound disagreement rapidly emerges as the null prediction saturates to a value of one, whist the empirical result grows exponentially.

Given the simplicity of the null model, it is perhaps not surprising that it does not capture the

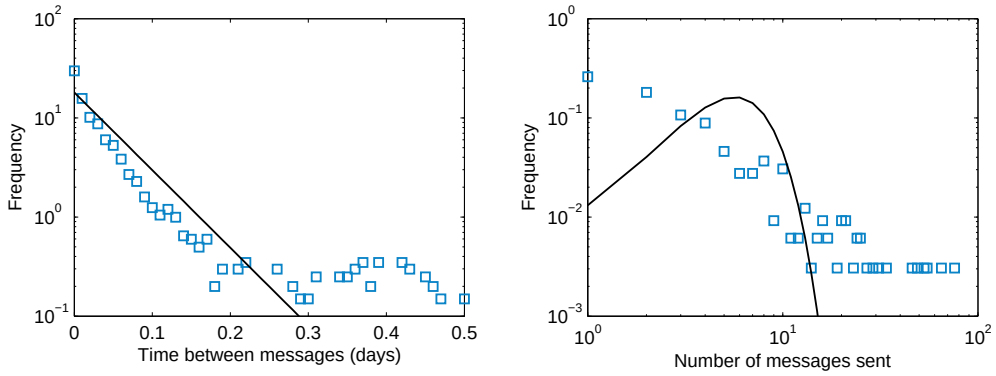

Figure 2: (Colour online) Comparison between message sending statistics in the data (blue squares) and null model (black lines).

behaviour of the real-world email network. What is not immediately clear, however, is precisely what features of the Kiel network have caused this large departure from the null prediction. We can begin to investigate this question by examining the effect of partially randomising the data to bring it closer in line with the assumptions of the null model.

We first check if the disagreement is rooted in the assumption that message times obey a Poisson process. The left panel of Figure 2 shows a histogram of waiting times between messages, compared with the exponential distribution assumed in the null model. The exponential distribution appears to provide a reasonable fit in short times, but misses the tail of the distribution. To see what effect this has on communicability, one approach is to recompute $\widetilde{\rho}$ for a dataset in which the original timestamps are preserved, but the message senders have been chosen at random from the whole network. The orange line in Figure 1 shows that this procedure results in a timeseries which fits the null prediction quite closely. Similarly, one can examine the effect of preserving the message senders and receivers but choosing the timestamps from a Poisson process with appropriate rate; the resulting timeseries for communicability lies very close to that of the original data (results of this experiment have been left out of Figure 1 to avoid clutter). Evidently correlations in the message times do not play a crucial role in the exponential growth of communicability.

The other key assumption of the null model was that message senders and receivers are chosen uniformly at random. In this case, there is no correlation between sending and receiving messages, and over a long time window the total number of messages sent or received by a given node should follow a Poisson distribution (much like the expected degree distribution in an Erdös-Renyi random graph). As shown in the right panel of Figure 2, this clearly does not hold for the Kiel data; some nodes are very much more active than others. To examine the effect of heterogeneous activity levels, we apply another randomisation process to the data. To generate the magenta line in Figure 1, message senders/receivers were chosen at random, but with likelihood proportional to the total number of messages sent/received in the whole data set. This process preserves the heterogeneity in node activity, but ignores any other structural

properties of the network. The resulting timeseries for $\widetilde{\rho}$ is a factor of 10 higher than the null model prediction, meaning this effect does play a role in the growth of communicability, however, it is still a long way below the result of the original data (note the logarithmic vertical axis in Figure 1). We can conclude that the bulk of the exponential growth is not accounted for purely by heterogeneity in node activity, and must therefore arise from other, as yet unknown, structural properties of network.

3 Pairwise interval graphs

3.1 Communicability in interval graphs

Recalling equation (2) from the introduction, communicability for temporal networks in continuous time is defined via the differential equation

$$S'(t) = -bS(t) - \big(I + S(t)\big) \log\big(I - aA(t)\big)\,. \tag{16}$$

Since the entries of an adjacency matrix are ones and zeros, it follows that $A(t)$ is a piecewise constant function of time. The interval graph formalism encodes temporal networks of this type as a list of quadruples $(i, j, t, \delta t)$ denoting a contact between nodes i and j lasting for the interval $[t, t + \delta t)$.

Let $t_1 < t_2 < \ldots$ denote the points of discontinuity for $A(t)$, so that $A(t) = A_n$ for $t \in [t_n, t_{n+1})$. These correspond to either the start or end of contact intervals. On each of these intervals (16) is a linear autonomous system, which can be straightforwardly solved to determine

$$S(t_{n+1}) = Q_n + \big(S(t_n) - Q_n\big) D_n\,, \tag{17}$$

where

$$\begin{aligned} Q_n &= b(b + L_n)^{-1} - I\,, \\ D_n &= e^{-(b+L_n)\Delta_n}\,, \\ L_n &= \log(I - aA_n)\,, \\ \Delta_n &= t_{n+1} - t_n\,. \end{aligned} \tag{18}$$

In this way the continuous time definition of communicability can be reduced to a recursion relation similar to its discrete time counterpart (1).

Further simplification is possible when the matrices $I - aA_n$ can be explicitly diagonalised. With application to telephone networks in mind, we consider the class of interval graphs composed of pairwise interactions representing one-to-one conversations. The calculation of the matrix logarithm decomposes over connected graph components, and for a single pair we have the formula

$$\log \begin{pmatrix} 1 & -a \\ -a & 1 \end{pmatrix} = \begin{pmatrix} \log\sqrt{1-a^2} & -\tanh^{-1}(a) \\ -\tanh^{-1}(a) & \log\sqrt{1-a^2} \end{pmatrix}\,. \tag{19}$$

Writing C_n for the collection of nodes engaged in pairwise contacts in the interval $[t_n, t_{n+1})$, we thus find

$$\big[L_n\big]_{ij} = -\tanh^{-1}(a)\big[A_n\big]_{ij} + \log\sqrt{1-a^2}\;\mathbb{I}_{\{i=j\in C_n\}}\,, \tag{20}$$

where $\mathbb{I}$ is the indicator function. Hence, introducing $c = b + \log(1-a)$, we have

$$\big[Q_n\big]_{ij} = \left(\frac{b}{2c} + \frac{b}{2(b+\log(1+a))}\right)\big[A_n\big]_{ij} + \left(\frac{b}{2c} - \frac{b}{2(b+\log(1+a))} - 1\right)\mathbb{I}_{\{i=j\in C_n\}}\,, \tag{21}$$

and

$$\begin{aligned}\big[D_n\big]_{ij} =& e^{-b\Delta_n}\frac{1}{2}\Big((1-a)^{-\Delta_n} + (1+a)^{-\Delta_n}\Big)\big[A_n\big]_{ij} \\ &+ e^{-b\Delta_n}\frac{1}{2}\Big((1-a)^{-\Delta_n} - (1+a)^{-\Delta_n}\Big)\mathbb{I}_{\{i=j\in C_n\}} \\ &+ e^{-b\Delta_n}\mathbb{I}_{\{i=j\notin C_n\}}\,.\end{aligned} \tag{22}$$

We now have a reasonably efficient procedure for computing the evolution of communicability in pairwise interval graphs: at each start or end point of an interval, we update Q_n and D_n according to the above formulas (which amounts to altering just four entries in each matrix) and then apply equation (17).

3.2 Null model

As previously, we are guided in the creation of the null model by the requirements that it should be uncorrelated and memoryless. Let us assume that contact intervals (or "calls", for convenience) occur as a Poisson random process with rate μ. Each call is between an independently randomly selected pair of individuals (with the condition that neither node is already in a call), and lasts for an exponentially distributed random time, with rate δ. We will write C_n for the set of nodes taking part in calls in a given time interval. Note that this null model differs from that in the previous section not just because contacts have a finite duration, but also because they are symmetric (undirected).

As before, we consider the time evolution of the average dynamic receive score. Summing equation (17) over columns, we find the update rule

$$\boldsymbol{r}_{n+1} = \boldsymbol{q}_n + D_n(\boldsymbol{r}_n - \boldsymbol{q}_n)\,, \tag{23}$$

where D_n is as in equation (22), and the vector $\boldsymbol{q}$ has entries $q_i = (b/c - 1)\mathbb{I}_{i\in C_n}$. To calculate the dynamical properties of the average $\rho(t) = \mathbb{E}(r_i(t))$, we consider the behaviour in a typical interval between the start of one call and the next.

Following the definition of the null model, the number of calls in progress at a given moment obeys an M/M/c/K queueing process (*i.e. a memoryless queue of length at most K with c servers; see [15] for details)*, with $c = K = \lfloor N/2 \rfloor$.

The equilibrium distribution for this process is

$$\mathbb{P}(|C_n| = 2k) \propto \frac{1}{k!}\left(\frac{\mu}{\delta}\right)^k\,. \tag{24}$$

After a short period in which the system moves away from its initial conditions, we will find that immediately after the start of the new call, there will be $k+1$ calls in total, where k has distribution (24). Moreover, at this moment the time before the next call, Γ, is distributed exponentially with rate μ. Due to the memoryless nature of the model, the $k+1$ calls have independently exponentially distributed durations Δ, which may or may not exceed Γ. For simplicity, we explore the possible cases separately.

If i and j have a conversation lasting for the whole of the interval $[t, t+\Gamma)$, then from equation (23) we have

$$\begin{aligned} r_i(t+\Gamma) = &\left(\frac{b}{c}-1\right)\left(1-e^{-c\Gamma}\right) \\ &+ \frac{e^{-b\Gamma}}{2}\Big((1-a)^{-\Gamma}+(1+a)^{-\Gamma}\Big)r_i(t) + \frac{e^{-b\Gamma}}{2}\Big((1-a)^{-\Gamma}-(1+a)^{-\Gamma}\Big)r_j(t)\,. \end{aligned} \tag{25}$$

Averaging first over the initial vector $\boldsymbol{r}(t)$ the symmetry of i and j gives

$$\mathbb{E}\big(r_i(t+\Gamma)\big|i\in C,\,\Delta>\Gamma\big) = \mathbb{E}\left(\left(\frac{b}{c}-1\right)\left(1-e^{-c\Gamma}\right)+e^{-c\Gamma}\rho(t)\right). \tag{26}$$

If the call finishes after a time $\Delta<\Gamma$, then we must apply equation (23) twice: first on the period $[t, t+\Delta)$ then again on $[t+\Delta, t+\Gamma)$. Averaging over the initial vector, the result is

$$\mathbb{E}\big(r_i(t+\Gamma)|i\in C,\Delta<\Gamma\big) = \mathbb{E}\left(e^{-b(\Gamma-\Delta)}\left[\left(\frac{b}{c}-1\right)\left(1-e^{-c\Delta}\right)+e^{-c\Delta}\rho(t)\right]\right). \tag{27}$$

Combining the two cases above, we average over the exponentially distributed times Γ and Δ:

$$\begin{aligned} &\mathbb{E}\big(\,r_i(t+\Gamma)|i\in C\big) \\ &\quad= \delta\mu\int_0^\infty\int_0^\Gamma \mathbb{E}\big(r_i(t+\Gamma)|\Delta<\Gamma\big)e^{-\delta\Delta-\mu\Gamma}\,d\Delta\,d\Gamma \\ &\qquad+\delta\mu\int_0^\infty\int_\Gamma^\infty \mathbb{E}\big(r_i(t+\Gamma)|\Delta>\Gamma\big)e^{-\delta\Delta-\mu\Gamma}\,d\Delta\,d\Gamma \\ &\quad= \frac{\mu(b+\mu+\delta)\big((b-c)+(\delta+\mu)\rho(t)\big)}{(b+\mu)(\delta+\mu)(\delta+\mu+c)} \end{aligned} \tag{28}$$

For the $N-2(k+1)$ nodes not involved in calls during $[t, t+\Gamma)$, we have

$$\mathbb{E}\big(r_i(t+\Gamma)|i\notin C\big) = \mu\int_0^\infty e^{-(b+\mu)\Gamma}\rho(t)\,d\Gamma = \frac{\mu}{b+\mu}\rho(t)\,. \tag{29}$$

Bringing all cases together, we reach

$$\begin{aligned} \mathbb{E}\big(r_i(t+\Gamma)\big) =&\,\mathbb{P}\big(i\in C\big)\mathbb{E}\big(\,r_i(t+\Gamma)|i\in C\big)+\mathbb{P}\big(i\notin C\big)\mathbb{E}\big(\,r_i(t+\Gamma)|i\notin C\big) \\ =&\frac{2(\mathbb{E}(k)+1)}{N}\frac{\mu(b+\mu+\delta)(b-c)}{(b+\mu)(\delta+\mu)(\delta+\mu+c)} \\ &+\left[\frac{2(\mathbb{E}(k)+1)}{N}\frac{\mu(b+\mu+\delta)}{(b+\mu)(\delta+\mu+c)}+\left(1-\frac{2(\mathbb{E}(k)+1)}{N}\right)\frac{\mu}{b+\mu}\right]\rho(t)\,. \end{aligned} \tag{30}$$

Now, for N large, the number of calls is approximately Poisson distributed with $\mathbb{E}(k) \approx \mu/\delta$, moreover, the average wait between call starts is $1/\mu$. We can thus write the following iterative equation for the dynamics of mean communicability $\rho(t)$, at time intervals of $1/\mu$:

$$\begin{aligned}\rho(t+1/\mu) = &\frac{2(\mu+\delta)}{\delta N}\frac{\mu(b+\mu+\delta)(b-c)}{(b+\mu)(\delta+\mu)(\delta+\mu+c)} \\ &+\left[\frac{2(\mu+\delta)}{\delta N}\frac{\mu(b+\mu+\delta)}{(b+\mu)(\delta+\mu+c)}+\left(1-\frac{2(\mu+\delta)}{\delta N}\right)\frac{\mu}{b+\mu}\right]\rho(t)\end{aligned} \tag{31}$$

This equation behaves in much the same way as the iteration (10) found previously. As time grows large, we have convergence to the equilibrium

$$\begin{aligned}\rho^* &= \frac{2\mu(b-c)(\delta+\mu+b)}{(\delta+\mu+c)b\delta N - 2\mu(b-c)(\delta+\mu)} \\ &\approx \frac{2\mu(b-c)}{b\delta N - 2\mu(b-c)}\,,\end{aligned} \tag{32}$$

where the second is the dominant contribution under the assumption that N and μ are both large.

This knowledge of the expected behaviour of the uncorrelated and memoryless null model allows for the calibration the communicability parameters. Putting $\rho^* = 1$ and solving for b we find

$$b \approx \frac{4\mu}{\delta N}\log\left(\frac{1}{1-a}\right)\,. \tag{33}$$

Once again, the mean gap μ^{-1} between call starts and the mean duration δ^{-1} of calls can easily be measured from data, or taken as rolling averages. Interestingly, for $\delta = 1$ and $a \ll 1$, (33) becomes $b \approx 4a\mu/N$. This expression is a factor of 2 larger than the calibration (14) found for directed contact sequences. It is possible that this difference is due to the present model being undirected where the other was directed; in effect each edge is counted twice.

3.3 Application to a phone network

The data gathered during the "Reality Mining" project [16] provides a useful test case for the dynamic communicability metric. A cohort of 106 test subjects were provided with mobile phones, and their usage was monitored over the course of the nine-month study. The call data may be naturally represented as an interval graph, where the interval $(i, j, t, \delta t)$ denotes a conversation between subjects i and j, which began at time t and had duration δt; a total of 2486 such calls were observed. The mean wait between calls was $\mu^{-1} = 11406s$ and mean call duration $\delta^{-1} = 69s$ (standard deviations $35429s$ and $133s$, respectively).

More conveniently for use with the iterative update rule, the data can be represented as a sequence of edge events (i_n, j_n, t_n), so that the adjacency matrix is encoded via

$$\left[A_n\right]_{ij} = \left[A_{n-1}\right]_{ij} \oplus \left(\delta_{i,i_n}\delta_{j,j_n} + \delta_{i,j_n}\delta_{j,i_n}\right)\,, \tag{34}$$

where $\oplus$ denotes the XOR operation, and A_0 is the zero matrix. In this prescription each interval (i, j, t, Δ) becomes a pair of edge events $n < m$ such that $i_n = i_m = i$, $j_n = j_m = j$, $t_n = t$ and $t_m = t + \Delta$. In this way the edge between i and j is switched on and off at the right time according to equation (34).

The thin blue line in Figure 3a shows the evolution of mean communicability over the course of the experiment, with fixed communicability parameters calibrated according equation (33), specifically $a = 5 \times 10^{-4}$ and $b = 2.8 \times 10^{-8}$ (a was chosen arbitrarily to provide clear figures for the example, and then b determined by the calibration). The initial communicability matrix was taken to have entries $S(0)_{ij} = (1 - \delta_{ij})/N$, so that all nodes are initially identical and the mean communicability is one. After an initial dip, there is an exponential explosion of weighted paths not predicted by the uncorrelated null model. As observed previously in the real-world contact sequence networks, this behaviour is intimately related to the structure of the network; when the call participants are randomised (thin orange line) exponential growth is not evident and communicability stays of order one.

Some characteristics of the Reality Mining network vary considerably over the course of the experiment; in particular, Figure 3b shows the large variation in the frequency of calls. One result of this variability is, for example, that what might be suitable calibration of communicability parameters on day 100 would be quite inappropriate for day 250. This problem can be avoided by calibrating parameters using a rolling average. The solid blue and orange lines in Figure 3a show the evolution of mean communicability in the real and randomised networks, with a fixed and b calculated according to (33) with μ and δ determined as an average over the previous 100 calls. By using dynamic calibration, the dips in communicability seen at the start and end of the experiment are removed. We can conclude that these dips were caused by changes in the temporal features of the network, rather than changes in the structure, which could not be compensated for by dynamic calibration of parameters.

4 Discussion

The purpose of this article has been to examine the behaviour of the communicability metrics for dynamic centrality using simple null models. Two classes of temporal networks were considered: contact sequences and interval graphs, with the corresponding definitions of communicability given by equations (1) and (2). In both cases, null models were created using the assumption that edges should be uncorrelated in their positions and times. The parameters of the null models are the number of nodes N, the average wait between contacts μ^{-1}, and for the interval graphs the average duration of contact δ^{-1}. Analysis of the null models revealed in both cases a simple trade-off between the edge attenuation and time downweighting parameters. Choosing to calibrate the metrics so that the mean dynamic receive score is one yields the relationship $b = 2a\mu/N$ for contact sequences and $b = -4\mu \log(1-a)/\delta N$ for interval graphs.

This analysis of the interaction between parameters removes some of the arbitrary nature of their definition and should cut down on the effort needed to find suitable values for a given

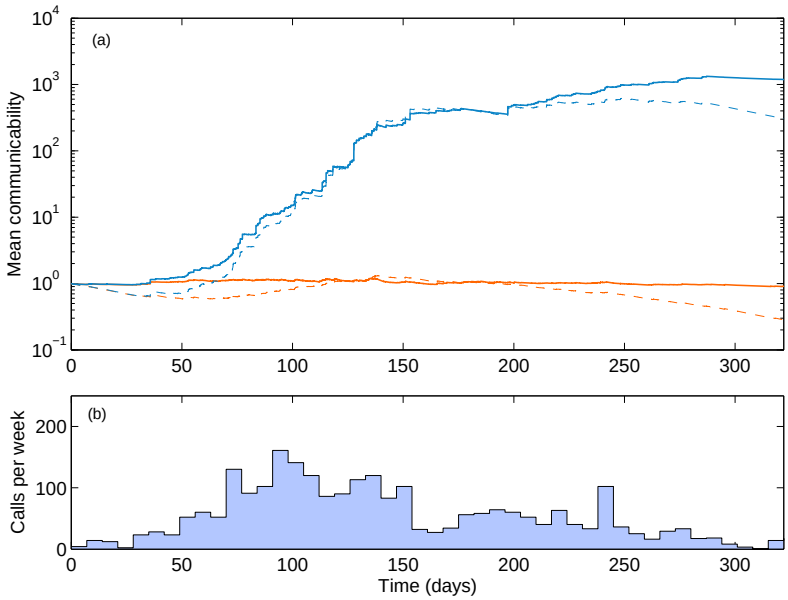

Figure 3: (Colour online) (a) Time series of mean communicability in the MIT phone network with edge-attenuation parameter $a = 5 \times 10^{-4}$. The true data (blue) is compared with randomised data (orange), with solid and dashed lines used to denote adaptive and fixed choices for the time-down-weighting parameter b, respectively. Ensemble variability is again negligible on this scale. (b) Number of calls made per week of the trial, used to calibrate the adaptive time downweighting.

dataset. Moreover, it is possible to adjust the parameters in a systematic way to compensate for changes in the level of activity of a network. This dynamic calibration was demonstrated for the MIT "Reality Mining" dataset, where the dip in communicability at the start and end of the experiment were shown to be simple artefacts of the overall level of activity, rather than a change in the underlying structure of the network – something which is not clear *a priori*. Dynamic parameter setting is potentially useful for de-trending or removing oscillations from data, however, it does raise questions over how to draw precise comparisons between results generated with different parameters.

As seen in Figures 1 and 3, there is a large dependency between the prediction of the null models and the empirical measurement of communicability in real-world telecommunications data. Where average dynamic receive scores in the null model converge to one, the empirical results are seen to grow exponentially. This finding suggests that, perhaps unsurprisingly, the real-world networks are organised to allow for much more efficient dissemination of information than is possible by random communication. A better understanding of the root causes of this phenomenon can be gained by examining communicability in partially randomised versions of the same data. For example, in both the Kiel and MIT networks randomising the location of edges in the network (but not the temporal data) leads to results in line with the null model, implying that temporal variations have only a secondary affect on the growth of communicability. Closer analysis of the Kiel data showed that heterogeneity in node activity was at least partly responsible, however, this is far from a complete explanation. Further theoretical work in this area could be very valuable; one possibility is to explore the role of 'flow motifs' [17], as repeated patterns of messages could accelerate the growth of communicability between certain nodes.

One consequence of the uncorrelated nature of the null model is that the expected dynamics of communicability of a single node is the same as that of the total communicability of the whole network. This fact was heavily exploited in the theoretical analysis presented here, and the numerical tests were all conducted at the whole-network level. The metrics discussed in this article have already found use in commercial contexts [18], and the applications of temporal network analysis are only like to grow in coming years. In these practical applications, the real utility of dynamic communicability is in distinguishing the roles of different nodes within the network. Just as comparison to null models provides useful insights into the dynamical behaviour of the network as a whole, more sophisticated future analyses (perhaps with more detailed null models) could extend these benefits to the study of individual node dynamics.

References

[1] Bruce R Lindsay. *Social Media and Disasters: Current Uses, Future Options, and Policy Considerations*, volume 41987. Congressional Research Service, 2011.

[2] Stanley Wasserman and Joseph Galaskiewicz. *Advances in social network analysis: Research in the social and behavioral sciences*. Sage, 1994.

[3] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18:39 – 43, 1953.

[4] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97 – 125, 2012.

[5] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia. In *Proceedings of the 3rd Workshop on Social Network Systems*, SNS '10, pages 31 – 36, New York, 2010. ACM.

[6] Kristina Lerman, Rumi Ghosh, and Jeon Hyung Kang. Centrality metric for dynamic networks. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, pages 70 – 77, New York, NY, USA, 2010. ACM.

[7] E. Estrada and D. Higham. Network properties revealed through matrix functions. *SIAM Review*, 52(4):696–714, 2010.

[8] Raj Kumar Pan and Jari Saramäki. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E*, 84:016105, 2011.

[9] Hyoungshick Kim and Ross Anderson. Temporal node centrality in complex networks. *Phys. Rev. E*, 85:026107, 2012.

[10] Hartmut H. K. Lentz, Thomas Selhorst, and Igor M. Sokolov. Unfolding accessibility provides a macroscopic approach to temporal networks. *Phys. Rev. Lett.*, 110:118701, 2013.

[11] Peter Grindrod, Mark C. Parsons, Desmond J. Higham, and Ernesto Estrada. Communicability across evolving networks. *Phys. Rev. E*, 83:046120, 2011.

[12] P. Grindrod and D. Higham. A matrix iteration for dynamic network summaries. *SIAM Review*, 55(1):118–128, 2013.

[13] Peter Grindrod and Desmond J. Higham. A dynamical systems view of network centrality. *Proc. R. Soc. A*, 470:2165, 2014 .

[14] Holger Ebel, Lutz Ingo Mielsch, and Stefan Bornholdt. Scale-free topology of e-mail networks. *Phys. Rev. E*, 66:035103, 2002.

[15] Leonard Kleinrock. *Queueing systems. volume 1: Theory.* Wiley-Interscience, 1975.

[16] Nathan Eagle, Alex Pentland, and David Lazer. Inferring social network structure using mobile phone data. *Proc. Nat. Ac. Sci.*, 106:15274–15278, 2009.

[17] Luis E. C. Rocha and Vincent D. Blondel. Flow motifs reveal limitations of the static framework to represent human interactions. *Phys. Rev. E*, 87:042814, 2013.

[18] Peter Grindrod, Desmond J Higham, and Peter Laflin. Collaboration blooms from SIAM news article. *SIAM News*, 45(10), 2012.